Relatório de CT-213:
Aprendizado por reforço livre de modelo

Henrique F. Feitosa

Instituto Tecnológico de Aeronáutica,
São José dos Campos, São Paulo, Brasil

1 Introdução

Nessa prática, buscou-se implementar os algoritmos de aprendizado por reforço livre de modello: Sarsa e Q-Learning. Para testar os algoritmos implementados, eles foram submetidos a um processo que os submetia a um MDP que consiste em um tabuleiro unidimensional em que as ações são "STOP", "LEFT"e "RIGHT", sendo que se a partícula realizar o comando "LEFT"na célula da extrema esquerda ela vai para a célula da extrema direita e vice-versa.

Após esse teste, treinou-se uma política para o robô seguidor de linha com os dois algoritmos e analisou-se o resultado.

2 Resultados e Discussão

Os resultados obtidos estão nos testes figuras 1 e 2.

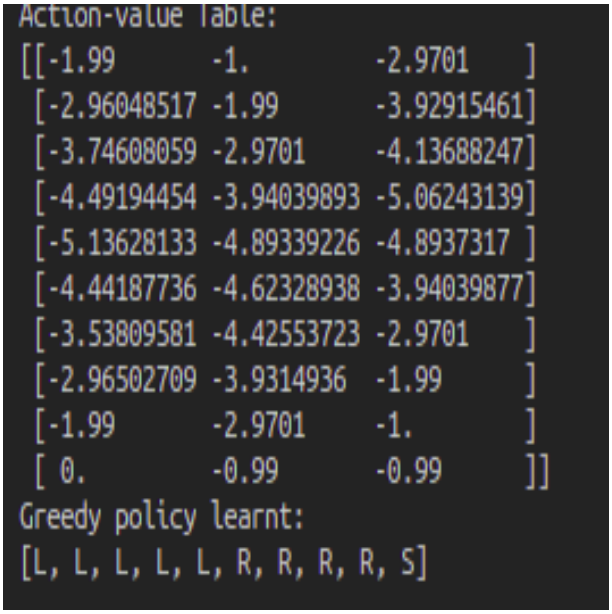

Figura 1. Mostra o resultado do algoritmo Q-learning

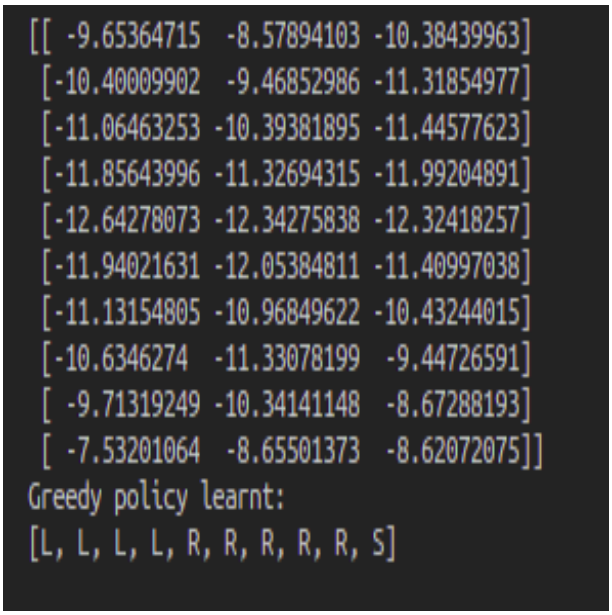

Figura 2. Mostra o resultado do algoritmo Sarsa

Assim, pode-se perceber que o Q-learning tem um pouco mais de exploration em comparação ao Sarsa, que é predominantemente explotation. Assim, é de se esperar que os resultados achados pelo Q-learning encontrem valores menores. Os resultados da política do robô seguidor de linha estão representados pelas figuras 3,4,5,6,7,8,9 e 10.

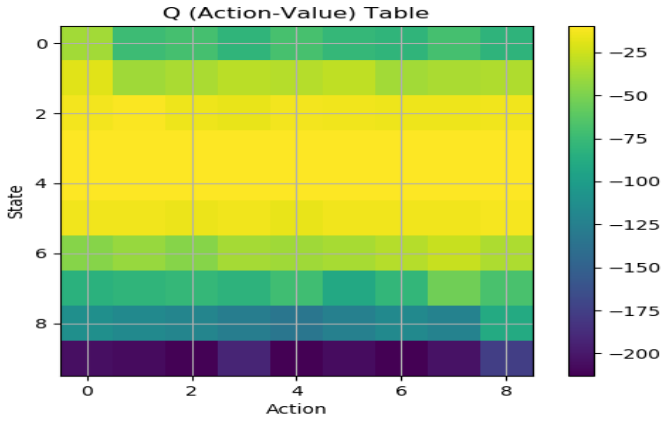

Figura 3. Mostra o os valores para cada estado e cada ação tomada no algoritmo Sarsa

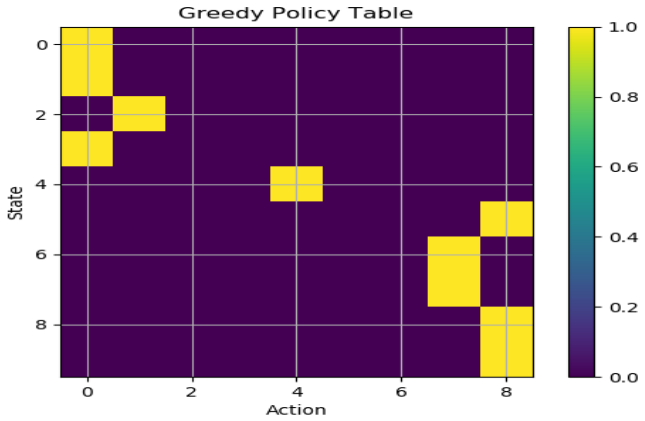

Figura 4. Mostra a política encontrada pelo para cada estado e a ação tomada pelo algoritmo Sarsa

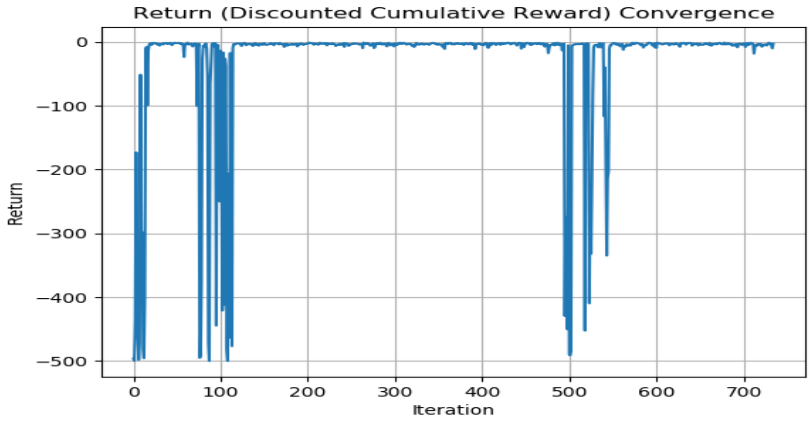

Figura 5. Mostra a convergência encontrada pelo algoritmo Sarsa ao longo das iterações

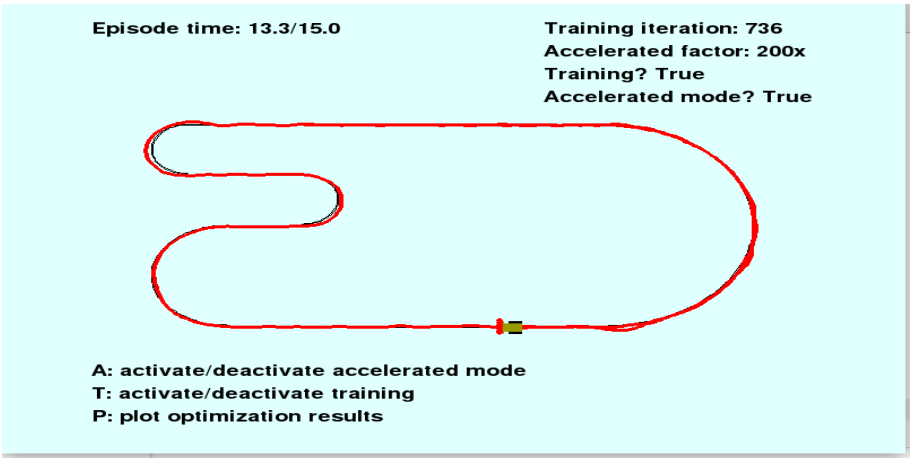

Figura 6. Mostra a política ótima encontrada do algoritmo Sarsa

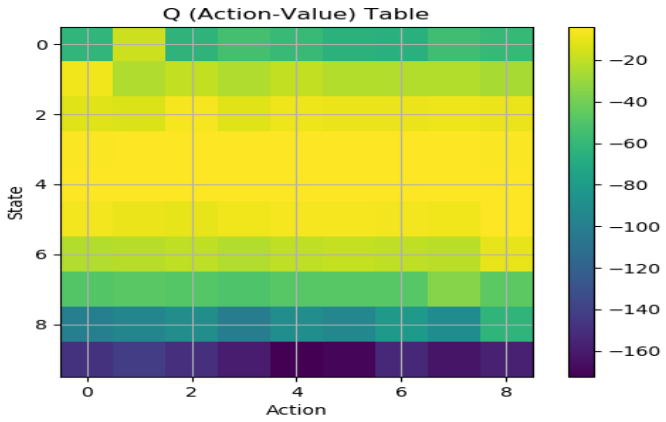

Figura 7. Mostra o os valores para cada estado e cada ação tomada no algoritmo Q-learning

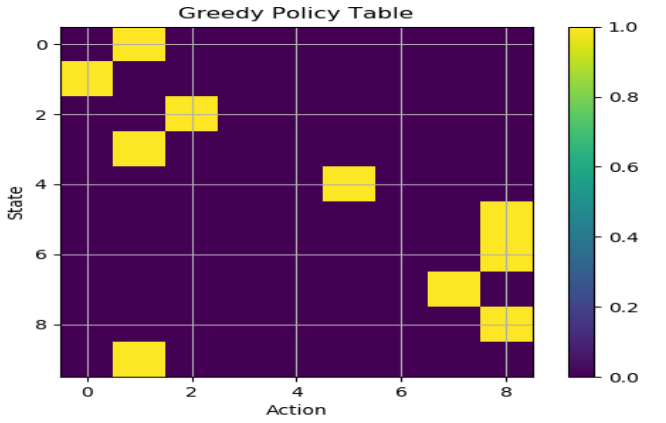

Figura 8. Mostra a política encontrada pelo para cada estado e a ação tomada pelo algoritmo Q-learning

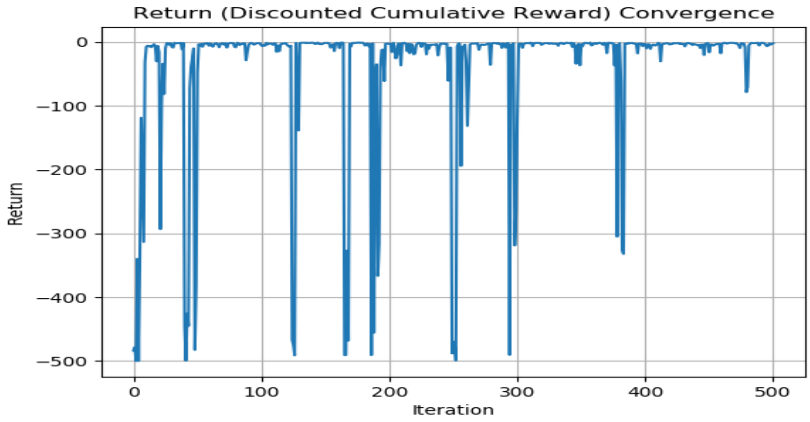

Figura 9. Mostra a convergência encontrada pelo algoritmo Q-lerning ao longo das iterações

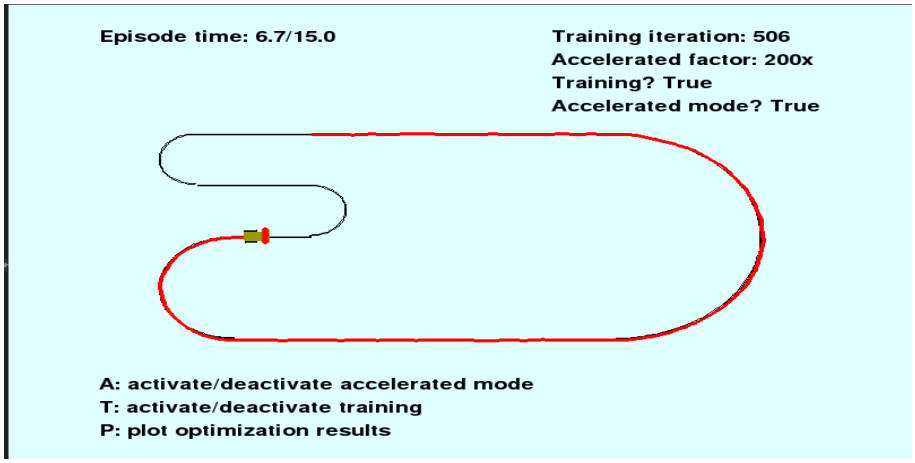

Figura 10. Mostra a política ótima encontrada do algoritmo Q-learning

Pelas figuras 5 e 9, pode-se perceber que a convergência dos parâmetros foi aceitável, pode-se confirmar isso pelas políticas mostradas nas figuras 10 e 6 que mostram o caminho encontrado pelo robô depois de um número considerável de informações. Além disso, é notável perceber que durante as iterações, o algoritmo Q-learning muitas vezes ia por caminhos não tão bons, visto que o robô saia da linha facilmente, isso confirma sua característica forte de exploration.